

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-114484

(43) Date of publication of application : 02.05.1997

G10L 3/00
G10L 3/00

(71)Applicant : ATR ONSEI HONYAKU TSUSHIN
KENKYUSHO:KK

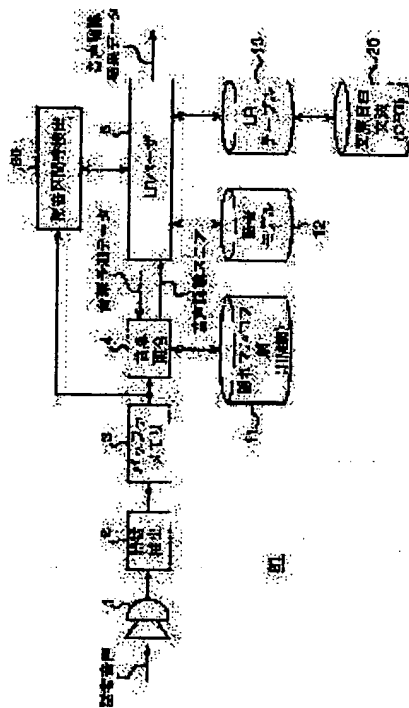
(72)Inventor : TAKEZAWA TOSHIYUKI
MORIMOTO TAKUMA

(54) VOICE RECOGNITION DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To eliminate the ambiguity of modification relation in a syntax analysis by determining the modification relation of the syntax analysis as to each voice section divided with a voiceless section, etc., and recognizing a spoken voice.

SOLUTION: A detection part 30 for a voiceless section, etc., detects the voiceless section, etc., including a section based upon a pause, a redundant word, rhythmical information, etc., according to a time series of feature parameters outputted from a buffer memory 3, and outputs its detection signal to an LR purser 5. The LR purser 5 reads in data in the speech section of a section unit indicated with the inputted detection signal and performs a section-limited HMM-LR process using an HMM-LR method for the speech section. Consequently, the modification relation in the speech section of each section unit is determined, and then modification relation between words, phrases, or clauses in the speech sections of different section units is determined.



LEGAL STATUS

[Date of request for examination] 24.10.1995

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 2880436

[Date of registration] 29.01.1999

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right] 29.01.2003

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平9-114484

(43)公開日 平成9年(1997)5月2日

(51) Int.Cl.⁸
G 1 0 L 3/00

識別記号 535
561

F I
G 1 0 L 3/00

技術表示箇所

5 3 5
5 6 1 G

審査請求 有 請求項の数4 OL (全 13 頁)

(21)出願番号 特願平7-275866

(22)出願日 平成7年(1995)10月24日

(71)出願人 593118597

株式会社エィ・ティ・アール音声翻訳通信
研究所
京都府相楽郡精華町大字乾谷小字三平谷5
番地

(72)發明者 竹澤 寿幸

京都府相楽郡精華町大字乾谷小字三平谷5番地 株式会社エィ・ティ・アール音声翻訳通信研究所内

(72)発明者 森元 逞

京都府相楽郡精華町大字乾谷小字三平谷5番地 株式会社エィ・ティ・アール音声翻訳通信研究所内

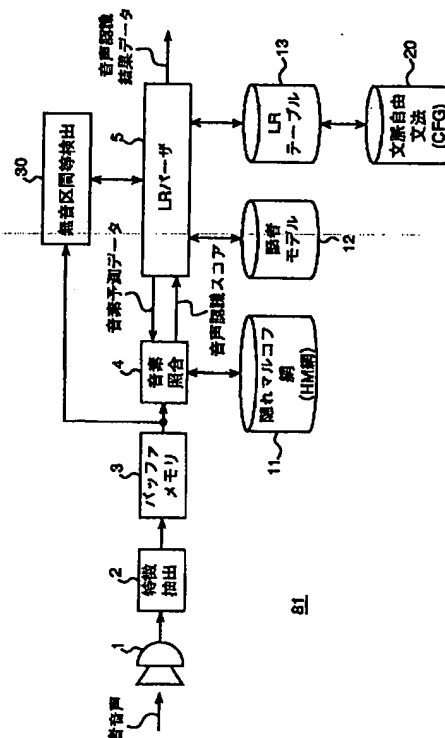
(74)代理人 弁理士 青山 葆 (外2名)

(54) 【発明の名称】 音声認識装置

(57) 【要約】

【課題】 統語解析における係り受け関係の曖昧性を解消することのできる音声認識装置を提供する。

【解決手段】 入力された発声音声を生声認識して音声認識結果を出力する音声認識手段を備えた音声認識装置において、入力された発声音声に基づいてポーズと冗長語と句又は節の境界とのうちの少なくとも1つを検出して検出信号を出力する検出手段を備え、音声認識手段は、ポーズと冗長語と句又は節の境界とのうちの少なくとも1つによって分割された複数の音声区間からなる入力された発声音声の各音声区間の音声認識処理をした後、異なる音声区間に属する語、句又は節の間の係り受け関係を決定して発声音声の音声認識をする。



【特許請求の範囲】

【請求項1】 入力された発声音声を音声認識して音声認識結果を出力する音声認識手段を備えた音声認識装置において、

入力された発声音声に基づいてポーズと冗長語と句又は節の境界とのうちの少なくとも1つを検出して検出信号を出力する検出手段を備え、

上記音声認識手段は、上記検出信号に基づいて統語解析における係り受け関係を決定して上記発声音声の音声認識をすることを特徴とする音声認識装置。

【請求項2】 上記音声認識手段は、上記ポーズと冗長語と句又は節の境界とのうちの少なくとも1つによって分割された複数の音声区間からなる入力された発声音声の各音声区間について音声認識処理をした後、異なる音声区間に属する語、句又は節の間の係り受け関係を決定して、上記入力された発声音声の音声認識をすることを特徴とする請求項1記載の音声認識装置。

【請求項3】 上記検出手段は、上記発声音声のパワーが、所定の時間の範囲だけ、所定のしきい値以下である第1の条件と、上記発声音声のゼロクロス数が、所定の時間の間において、所定のしきい値以上である第2の条件とのうち少なくとも1つの条件が満足することを検出することにより上記ポーズを検出することを特徴とする請求項1又は2記載の音声認識装置。

【請求項4】 上記検出手段は、上記ポーズと冗長語と句又は節の境界とのうちの少なくとも1つを、それぞれの予め決められた言語モデルに一致するか否かを判断することにより検出することを特徴とする請求項1又は2記載の音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は音声認識装置に関し、特に、発声音声中におけるポーズ（無音区間）又は冗長語などの無音区間等を検出して連続的に音声認識を実行する音声認識装置に関する。なお、本明細書では、ポーズと冗長語並びに韻律的な情報等を手がかりとする区切りとを含むものを無音区間等という。

【0002】

【従来の技術】 近年、連続音声認識の研究が盛んに行われ、いくつかの研究機関で文音声認識システムが構築されている。これらのシステムの多くは丁寧に発声された音声を入力対象にしている。しかしながら、人間同士のコミュニケーションでは、「あー」、「えー」となどに代表される冗長語や、一時的に発声音声が無い無音区間等の状態のポーズである言い淀みや言い誤り及び言い直しなどが頻繁に出現する。

【0003】 図9は、図2に示す例文「きれいな黒い髪の女の子を見た」を従来例の連続音声認識装置で音声認識処理を実行するときの音声認識動作をスタック形式で示す図である。従来例の連続音声認識装置の音声認識動

作について図9を参照して説明する。まず、図9の状態スタック201に示すように、「きれいな」という発声音声の系列が認識されて文字として積まれる。次に、状態スタック201における「きれいな」という文字は音声認識用辞書に載っているため、状態スタック202に示すように形容詞句を表す「adj」という文字に変換される。次に、「黒い」という発声音声の系列が認識されて状態スタック203に示すように文字として積まれ、状態スタック203における「黒い」は音声認識辞書に載っているため状態スタック204に示すように形容詞句を表す「adj」という文字に変換される。

【0004】 次に、「髪の」という発声音声の系列が認識されて状態スタック205に示すようにさらに文字として積まれ、状態スタック205における「髪の」という文字は音声認識辞書に載っているため状態スタック206に示すように名詞句を表す「NP」という文字に変換される。さらに、状態スタック206において、形容詞句の「adj」と名詞句の「NP」とは名詞句の「NP」になるという構文規則が適用されて、「黒い」が変換された形容詞句の「adj」と「髪の」が変換された名詞句の「NP」とは、状態スタック207に示すように名詞句の「NP」に変換される。すなわち、状態スタック207における名詞句の「NP」は「黒い髪の」を表す。

【0005】 ここで、状態スタック207において、形容詞句の「adj」と名詞句の「NP」とは名詞句の「NP」になるという構文規則を適用するかしないか2つの選択枝がある。ここで、構文規則を適用すると「きれいな」は「髪の」に係ることになり、構文規則を適用しないと「きれいな」は「髪の」に係らない構文構造のまま係り受け関係の決定は以降の音声認識処理に持ち越されることになる。従って、このような場合、従来例の連続音声認識装置では、文字を積む装置を2つに分離して以降の音声認識を実行する。すなわち、一方の装置は、状態スタック207に構文規則を適用した状態スタック208に示す状態で以降の音声認識処理を実行し、他方の装置は、状態スタック207のままの状態状態で以降の音声認識処理を実行する。ここで、一方の装置の状態スタック208における名詞句の「NP」は「きれいな黒い髪の」を表す。

【0006】 一方の装置において、状態スタック209に示すように「きれいな黒い髪の」を表示する名詞句「NP」の上に、「女の子を」という発声音声の系列が認識されて文字として積まれ、状態スタック209における「女の子を」の文字は音声認識辞書に載っているため状態スタック210に示すように名詞句を表す「NP」という文字に変換される。次に状態スタック210において、名詞句の「NP」と名詞句の「NP」は名詞句の「NP」になるという構文規則が適用されて、状態スタック211の「きれいな黒い髪の女の子を」が表す名詞句の「NP」とは、状態スタック212に示すように名詞句の「NP」になるという構文規則が適用されて、状態スタック213の「きれいな黒い髪の女の子を見た」が表す名詞句の「NP」となる。

単位の音声区間のデータを読み込んで、当該音声区間に対してHMM-LR法を用いた区間制限付きHMM-LR処理を実行し、最後の区切り単位の末端まで到達すると入力された発声音声の1つの文に対して区間制限無しHMM-LR処理を実行することにより音声認識結果データを出力することを特徴とする。

【0015】ここで、上記SSSにおいては、音素の特徴空間上に割り当てられた確率的定常信号源(状態)の間の確率的な遷移により音声パラメータの時間的な推移を表現した確率モデルに対して、尤度最大化の基準に基づいて個々の状態をコンテキスト方向又は時間方向へ分割するという操作を繰り返すことによって、モデルの精密化を逐次的に実行する。

【0016】図1において、話者の発声音声はマイクロホン1に入力されて音声信号に変換された後、特徴抽出部2に入力される。特徴抽出部2は、入力された音声信号をA/D変換した後、例えばLPC分析を実行し、対数パワー、16次ケプストラム係数、 Δ 対数パワー及び16次 Δ ケプストラム係数を含む34次元の特徴パラメータを抽出する。抽出された特徴パラメータの時系列は

バッファメモリ3を介して音素照合部4に入力される。【0017】音素照合部4に接続されるHM網メモリ11内のHM網は、各状態をノードとする複数のネットワークとして表され、各状態はそれぞれ以下の情報を有する。

- (a) 状態番号
- (b) 受理可能なコンテキストクラス
- (c) 先行状態、及び後続状態のリスト
- (d) 出力確率密度分布のパラメータ
- (e) 自己遷移確率及び後続状態への遷移確率

【0018】なお、第1の実施形態において、HM網は、各分布がどの話者に由来するかを特定するため、所定の話者混合HM網を変換して作成する。ここで、出力確率密度関数は34次元の対角共分散行列をもつ混合ガウス分布であり、各分布はある特定の話者のサンプルを用いて学習されている。

【0019】音素照合部4は、LRパーザ5からの音素照合要求に応じて音素照合処理を実行する。このときに、LRパーザ5からは、音素照合区間及び照合対象音素とその前後の音素から成る音素コンテキスト情報が渡される。音素照合部4は、受け取った音素コンテキスト情報に基づいてそのようなコンテキストを受理することができるHM網上の状態を、先行状態リストと後続状態リストの制約内で連結することによって、1つのモデルが選択される。そして、このモデルを用いて音素照合区間内のデータに対する尤度が計算され、この尤度の値が音素照合スコアとしてLRパーザ5に返される。このときに用いられるモデルは、HMMと等価であるために、尤度の計算には通常のHMMで用いられている前向きパ

【0020】一方、無音区間等検出部30は、バッファメモリ3から出力される特徴パラメータの時系列に基づいてポーズや冗長語並びに韻律的な情報等を手がかりとする区切りを含む無音区間等を検出して、その検出信号をLRパーザ5に出力する。ここで、無音区間等検出部30は、冗長語については予め内部メモリに格納された冗長語の音素モデルと比較照合することにより冗長語として認識する一方、無音区間であるポーズについては以下の2つの条件のうちの1つが満足するときにポーズとして検出する。

(第1の検出条件) パワーが所定のしきい値レベル以下である時間 t_0 が例えば以下の範囲の値のとき。好ましくは、 $50 \text{ ミリ秒} \leq t_0 \leq 3 \text{ 秒}$ 。より好ましくは、 $50 \text{ ミリ秒} \leq t_0 \leq 500 \text{ ミリ秒}$ 。

(第2の検出条件) 入力された音声信号がゼロ電位と交差するゼロクロス数が所定のしきい値以上である時間 t_1 が例えば以下の範囲の値のとき。好ましくは、 $50 \text{ ミリ秒} \leq t_1 \leq 3 \text{ 秒}$ 。より好ましくは、 $50 \text{ ミリ秒} \leq t_1 \leq 500 \text{ ミリ秒}$ 。さらに、韻律的な情報等を手がかりとする区切りとは、具体的には、イントネーションが急激に上昇又は下降するときは、句又は節の境界であると推測される。これについては、入力される特徴パラメータのうち基本周波数が所定の傾斜の度合い以上で急激に上昇し又は下降して変化したことを検出することにより当該区切り又は境界と判別する。

【0021】そして、LRパーザ5は、無音区間等検出部30から入力された検出信号で示された区切り単位の音声区間のデータを読み込んで、当該音声区間に対してHMM-LR法を用いた区間制限付きHMM-LR処理を実行し、最後の区切り単位の末端まで到達すると入力された発声音声の1つの文に対して区間制限無しHMM-LR処理を実行することにより音声認識結果データを出力する。ここで、区間制限付きHMM-LR処理とは、1つの区切り単位の音声区間内に限って実行するHMMを用いたLRパーザ5による音声認識処理のことであり、区間制限無しHMM-LR処理とは、区間を限定せず、入力された発声音声の1つの文に対して、異なる区切り単位の音声区間に属する語、句又は節にLRテーブルメモリ13内の構文規則を適用して実行するHMMを用いたLRパーザ5による音声認識処理のことである。ここで、音声区間とは図5に示すように入力された発声音声の1つの文のうちの無音区間等(図5においては括弧を付して示している。)によって分割された1つの区間のことをいい、区切り単位とは図5において括弧を付して示すように音声区間と当該音声区間の後にある無音区間等とからなる1単位のことをいう。また、本明細書において、無音区間等とはポーズと冗長語並びに韻律的な情報等を手がかりとする区切りとを含むものをいい、ポーズ単位とは図5に示すようにポーズによって分

【0022】文脈自由文法データベースメモリ20内の所定の文脈自由文法(CFG)は公知の通り予め自動的に変換されてLRテーブルを作成してLRテーブルメモリ13に格納される。LRパーザ5は、例えば音素継続時間長モデルを含む話者モデルメモリ12と上記LRテーブルとを参照して、入力された音素予測データについて左から右方向に、後戻りなしに処理する。構文的にあいまいさがある場合は、スタックを分割してすべての候補の解析が平行して処理される。LRパーザ5は、LRテーブルメモリ13内のLRテーブルから次にくる音素を予測して音素予測データを音素照合部4に出力する。これに応答して、音素照合部4は、その音素に対応するHM網メモリ11内の情報を参照して照合し、その尤度を音声認識スコアとしてLRパーザ5に戻し、順次音素を接続していくことにより、連続音声の認識を行っている。

【0023】以上のように構成された第1の実施形態の連続音声認識装置81において、特徴抽出部2と音素照合部4とLRパーザ5とは、例えばデジタル電子計算機で構成される。

【0024】図6は、図1の連続音声認識装置81のLRパーザ5において実行される音声認識処理を示すフローチャートである。以下、図6を参照して音声認識処理について説明する。

【0025】図6に示すように、ステップS1においては、HMM作業域の初期化、並びにLRパーザ5の初期化を実行する。具体的には、状態スタック0のセルを1個作成する。ここで、連続音声認識装置81において用いるセルは、従来のHMM-LR法の音声認識の解析に必要な情報を保持するデータ構造、すなわち状態スタックを有するLR作業域と、音声認識スコアと確率テーブルとからなるHMM作業域とを有する。

【0026】そして、ステップS2において、無音区間等検出部30から入力された検出信号で示された区切り単位の音声区間のデータを読み込む。さらに、ステップS3において、音声データが読み込まれた区切り単位の音声区間に対してHMM-LR法を用いた区間制限付きHMM-LR処理を実行する。ステップS4において、複数の区切り単位のうち最後の区切り単位の末端まで到達したか否かが判断され、最後の区切り単位の末端まで到達していないときは(ステップS4においてNO)ステップS2に進み、ステップS2、S3の処理を繰り返す。一方、ステップS4において、最後の区切り単位の末端まで到達しているときは(ステップS4においてYES)ステップS5に進み、区間制限無しHMM-LR処理を実行して音声認識処理を終了する。

【0027】次に、図1の第1の実施形態の連続音声認識装置81の音声認識動作を図2に示す例文を用いて説明する。図2は、文の構造解析すなわち統語解析における係り受け関係の曖昧性を示す例文である。図2の例

文を文字列のみを認識して解析しようすると、図2の例文の上に矢印で示した第1の係り受け関係と例文の下に矢印で示した第2の係り受け関係の少なくとも2つの係り受け関係の曖昧性が残る。すなわち、「きれいな」が「女の子」に係る第1の係り受け関係の「きれいな女の子」であるのか、「きれいな」が「髪」に係る第2の係り受け関係の「きれいな髪」であるのかが不明である。本発明者らは、無音区間であるポーズを利用することにより上述の2つの係り受け関係のうちのいずれか1つに決定できることを見いだした。すなわち、「きれいな」と「黒い」との間に無音区間であるポーズ(図2においては、「きれいな」と「黒い」との間に「△」で示している。)があれば、「きれいな」が「女の子」に係る第1の係り受け関係であると決定でき、「髪の」と「女の子を」との間にポーズ(図2においては、「髪の」と「女の子を」との間に「△」で示している。)があれば、「きれいな」が「髪」に係る第2の係り受け関係であると決定できる。本発明は上述のポーズと係り受け関係との間の規則を利用して、統語解析における係り受け関係の曖昧性を取り除いて音声認識処理を実行している。

【0028】図3は、図2の例文において第1の係り受け関係を有する場合の連続音声認識装置81の音声認識動作をスタック形式で示す図である。以下に第1の係り受け関係を有する場合の音声認識動作を図3を参照して説明する。まず、図3の状態スタック51に示すように、LRパーザ5で「きれいな」という発声音声の系列が認識されて文字として積まれ、次に「きれいな」の認識処理の直後でポーズが無音区間等検出部30によって検出されて、検出信号が当該検出部30からLRパーザ5に入力されて「きれいな」という文字の上にポーズを表示する「△」として積まれる。次に、状態スタック51における「きれいな」という文字は音声認識用辞書に載っているので、状態スタック52に示すように形容詞句を表す「adj」という文字に変換される。次に、LRパーザ5で「黒い」という発声音声の系列が認識されて状態スタック53に示すようにポーズを表示する「△」の上に文字として積まれ、状態スタック53における「黒い」は音声認識辞書に載っているので状態スタック54に示すように形容詞句を表す「adj」という文字に変換される。ここで、状態スタック54において「きれいな」が変換された形容詞句の「adj」と「黒い」が変換された形容詞句の「adj」とには、間にポーズを表示する「△」が積まれているので構文規則は適用されない。

【0029】次に、LRパーザ5で「髪の」という発声音声の系列が認識されて状態スタック55に示すように「黒い」が変換された形容詞句の「adj」の上に文字として積まれ、状態スタック55における「髪の」という文字は音声認識辞書に載っているので状態スタック56に示すように名詞句を表す「n」に変換される。

6に示すように名詞句を表す「NP」という文字に変換される。さらに、状態スタック56において、形容詞句の「adj」と名詞句の「NP」とは名詞句の「NP」になるという構文規則が適用されて、「黒い」が変換された形容詞句の「adj」と「髪の毛」が変換された名詞句の「NP」とは状態スタック57に示すように名詞句の「NP」に変換される。すなわち、状態スタック57における名詞句の「NP」は「黒い髪の毛」を表す。次に、「女の子を」という発声音声の系列が認識されて状態スタック58に示すように「黒い髪の毛」を表す名詞句の「NP」の上に文字として積まれ、状態スタック58における「女の子を」の文字は音声認識辞書に載っているため状態スタック59に示すように名詞句を表す「NP」という文字に変換される。

【0030】次に状態スタック59において、名詞句の「NP」と名詞句の「NP」は名詞句の「NP」になるという構文規則が適用されて、状態スタック59の「黒い髪の毛」が変換された名詞句の「NP」と「女の子を」が変換された名詞句の「NP」は状態スタック60に示すように名詞句の「NP」に変換される。ここで、状態スタック60の名詞句の「NP」は「黒い髪の毛の女の子」を表す。そして、LRパーザ5で「見た」という発声音声の系列が認識されて状態スタック61に示すように「黒い髪の毛の女の子」を表す名詞句の「NP」の上に文字として積まれ、状態スタック61における「見た」は音声認識用辞書に載っているため状態スタック62に示すように動詞句を表す「VP」に変換される。

【0031】そして、最後のポーズ単位の末端まで到達していると判断されて、ポーズを表示する「△」の前後に位置する「きれいな」を表す形容詞句の「adj」と「黒い髪の毛の女の子を」を表す名詞句の「NP」とに、形容詞句の「adj」と名詞句の「NP」とは名詞句の「NP」になるという構文規則が適用されて状態スタック63に示すように名詞句の「NP」に変換される。ここで、状態スタック63の名詞句の「NP」は、「きれいな」が「女の子を」に係る構造の「きれいな黒い髪の毛の女の子を」を表す。さらに、状態スタック64に示すように文章を表す「S」に変換されて、「きれいな」が「女の子を」に係るような構造の音声認識結果のみが出力される。以上のようにポーズを表示する「△」の前後に位置する「きれいな」を表す形容詞句の「adj」と「黒い髪の毛の女の子を」を表す名詞句の「NP」との間の構文規則の適用を最後のポーズ単位の末端まで到達してから実行するので、「きれいな」が「女の子を」に係る構造の音声認識結果のみを出力することができる。以上のように第1の実施形態では、複数の音声区間からなる入力された発声音声の1つの文の各音声区間の音声認識を実行した後、区間を限定せず、入力された発声音声の1つの文に対して異なる音声区間に属する語、句又は節の間にも、ポーズを挿入して構文規則を適用して

異なる音声区間に属する語、句又は節の間の係り受け関係を決定している。

【0032】図4は、図2の例文において第2の係り受け関係を有する場合の連続音声認識装置81の音声認識動作をスタック形式で示す図である。以下に第2の係り受け関係を有する場合の音声認識動作を図4を参照して説明する。まず、図4の状態スタック151に示すように、LRパーザ5で「きれいな」という発声音声の系列が認識されて文字として積まれる。次に、状態スタック151における「きれいな」という文字は音声認識用辞書に載っているため、状態スタック152に示すように形容詞句を表す「adj」という文字に変換される。次に、LRパーザ5で「黒い」という発声音声の系列が認識されて状態スタック153に示すように「きれいな」を表す形容詞句の「adj」の上に文字として積まれ、状態スタック153における「黒い」は音声認識辞書に載っているため状態スタック154に示すように形容詞句を表す「adj」という文字に変換される。

【0033】次に、LRパーザ5で「髪の毛」という発声音声の系列が認識されて状態スタック155に示すように「黒い」を表す形容詞句の「adj」の上に文字として積まれ、次に「髪の毛」の認識処理の直後でポーズが無音区間等検出部30によって検出されて、検出信号が当該検出部30からLRパーザ5に入力されてポーズを表示する「△」として「髪の毛」の文字の上に積まれる。そして、状態スタック155における「髪の毛」という文字は音声認識辞書に載っているため状態スタック156に示すように名詞句を表す「NP」という文字に変換される。

【0034】さらに、状態スタック156において、形容詞句の「adj」と名詞句の「NP」とは名詞句の「NP」になるという構文規則が適用されて、「黒い」が変換された形容詞句の「adj」と「髪の毛」が変換された名詞句の「NP」とは状態スタック157に示すように名詞句の「NP」に変換される。すなわち、状態スタック157における名詞句の「NP」は「黒い髪の毛」を表す。さらに、状態スタック157において、形容詞句の「adj」と名詞句の「NP」とは名詞句の「NP」になるという構文規則が適用されて、「きれいな」が変換された形容詞句の「adj」と「黒い髪の毛」を表す名詞句の「NP」とは状態スタック158に示すように名詞句の「NP」に変換される。これによって、「きれいな」が「髪の毛」にかかる構造として認識される。

【0035】次に、「女の子を」という発声音声の系列が認識されて状態スタック159に示すようにポーズを表示する「△」の上に文字として積まれ、状態スタック159における「女の子を」の文字は音声認識辞書に載っているため状態スタック160に示すように名詞句を表す「NP」という文字に変換される。ここで、状態スタック160において「きれいな黒い髪の毛」を表す名詞

句の「NP」と「女の子を」が変換された名詞句の「NP」とには、間にポーズを表示する「△」が積まれているので構文規則は適用されない。そして、「見た」という発声音声の系列が認識されて状態スタック161に示すように「女の子を」が変換された名詞句の「NP」の上に文字として積まれ、状態スタック161における「見た」は音声認識用辞書に載っているため状態スタック162に示すように動詞句を表す「VP」に変換される。

【0036】そして、LRパーザ5で最後のポーズ単位の末端まで到達していると判断されて、状態スタック162におけるポーズを表示する「△」の前後に位置する「きれいな黒い髪の」を表す名詞句の「NP」と「女の子を」を表す名詞句の「NP」とにLRテーブルメモリ13内の構文規則が適用されて、状態スタック163に示すように名詞句の「NP」に変換されて、さらに状態スタック164に示すように文章を表す「S」の文字に変換されて、「きれいな」が「黒い髪の」に係るような構造の音声認識結果のみが出力される。

【0037】以上の第1の実施形態の連続音声認識装置81は、無音区間等を検出して検出信号を出力する無音区間等検出部30を備え、LRパーザ5は、無音区間等検出部30から入力された検出信号で示された区切り単位の音声区間のデータを読み込んで、当該音声区間に対してHMM-LR法を用いた区間制限付きHMM-LR処理を実行し、最後の区切り単位の末端まで到達すると入力された発声音声の1つの文に対して区間制限無しHMM-LR処理を実行することにより音声認識結果データを出力する。これによって、各区切り単位の音声区間内における係り受け関係を決定した後、異なる区切り単位の音声区間に属する語、句又は節の間の係り受け関係を決定できるので、統語解析における係り受け関係の曖昧性を解消することができる。

【0038】＜第2の実施形態＞図7は、本発明に係る第2の実施形態である連続音声認識装置82のブロック図である。図7の第2の実施形態の連続音声認識装置82は、図1の第1の実施形態の連続音声認識装置81の隠れマルコフ網メモリ11に代えて隠れマルコフ網メモリ11aを備え、かつ無音区間等検出部30を除いて構成される。第2の実施形態の連続音声認識装置82においては、ポーズや冗長語並びに韻律的な情報等を手がかりとする区切りなどの無音区間等をHMMでモデル化したモデルが隠れマルコフ網メモリ11aに格納され、当該モデルを用いて無音区間等の検出を音素照合部4で行っている。

【0039】図8は、図7の連続音声認識装置82において実行される音声認識処理を示すフローチャートである。以下、図8を参照して第2の実施形態の連続音声認識装置82の音声認識処理について説明する。まず、ス

にLRパーザ5の初期化を実行する。具体的には、状態スタック0のセルを1個作成する。そして、ステップS11において、例えば、特徴パラメータの処理単位である音声フレーム（例えば20ミリ秒）毎に音声データの読み込みを行い、ステップS12において区間制限付きHMM-LR処理を実行する。次にステップS13において無音区間等を検出したか否かが判断され、無音区間等を検出していない場合はステップS11に進みステップS11、S12の処理が繰り返され、無音区間等を検出した場合はステップS14に進む。

【0040】ステップS14において、すべての音声区間の音声認識処理が終了したか否かが判断され、すべての音声区間の処理が終了していないときは（ステップS14においてNO）ステップS11に進み、ステップS11、S12、S13の処理を繰り返し、すべての音声区間の処理が終了したと判断されると（ステップS14においてYES）ステップS15に進み、入力された発声音声の1つの文に対して区間制限無しHMM-LR処理を実行して音声認識処理を終了する。

【0041】以上の第2の実施形態の連続音声認識装置82は、無音区間等の検出を隠れマルコフ網メモリ11aに格納されたHMMでモデル化した無音区間等のモデルを使用して音素照合部4で行い、LRパーザ5は、音声データを読み込んで、1つの音声区間に対してHMM-LR法を用いた区間制限付きHMM-LR処理を実行し、各音声区間についての処理が終了すると入力された発声音声の1つの文に対して区間制限無しHMM-LR処理を実行することにより音声認識結果データを出力する。これによって、各区切り単位の音声区間内における係り受け関係を決定した後、異なる音声区間に属する語、句又は節の間の係り受け関係を決定できるので、統語解析における係り受け関係の曖昧性を解消することができる。

【0042】以上の第1と第2の実施形態においては、入力された発声音声の1つの文に対して区間制限無しHMM-LR処理を実行することにより音声認識結果データを出力するようにした。しかしながら、本発明はこれに限らず、入力された発声音声の1つの句又は節等の1つのシーケンスの発声音声に対して区間制限無しHMM-LR処理を実行するようにしてもよいし、連続音声認識装置のスイッチがオンされてからオフされるまでの間に入力される発声音声に対して区間制限無しHMM-LR処理を実行するようにしてもよい。以上のように構成しても第1と第2の実施形態と同様に動作し同様の効果を有する。

【0043】以上の第1と第2の実施形態においては、HMM-LR法を用いた音声認識装置について述べているが、本発明はこれに限らず、ニューラルネットワークを用いた音声認識装置など他の種類の音声認識装置に適

【0044】

【発明の効果】本発明に係る請求項1記載の音声認識装置は、入力された発声音声に基づいてポーズと冗長語と句又は節の境界とのうちの少なくとも1つを検出して検出信号を出力する検出手段を備え、上記音声認識手段は、上記検出信号に基づいて統語解析における係り受け関係を決定して上記発声音声の音声認識をしている。これによって、統語解析における係り受け関係の曖昧性を解消できる。

【0045】また、請求項2記載の音声認識装置は、請求項1記載の音声認識装置において、上記音声認識手段は、上記ポーズと冗長語と句又は節の境界とのうちの少なくとも1つによって分割された複数の音声区間からなる入力された発声音声の各音声区間について音声認識処理をした後、異なる音声区間に属する語、句又は節の間の係り受け関係を決定して、上記入力された発声音声の音声認識をしている。これによって、統語解析における係り受け関係の曖昧性を解消できる。

【0046】さらに、請求項3記載の音声認識装置は、請求項1又は2記載の音声認識装置において、上記検出手段は、上記発声音声のパワーが、所定の時間の範囲だけ、所定のしきい値以下である第1の条件と、上記発声音声のゼロクロス数が、所定の時間の間において、所定のしきい値以上である第2の条件とのうち少なくとも1つの条件が満足することを検出することにより上記ポーズを検出している。これによって、上記ポーズに基づいて統語解析における係り受け関係を決定でき、統語解析における係り受け関係の曖昧性を解消できる。

【0047】またさらに、請求項4記載の音声認識装置は、請求項1又は2記載の音声認識装置において、上記検出手段は、上記ポーズと冗長語と句又は節の境界とのうちの少なくとも1つを、それぞれの予め決められた言語モデルに一致するか否かを判断することにより検出している。これによって、音声認識過程で上記ポーズと冗長語と句又は節の境界とのうちの少なくとも1つを検出

でき、統語解析における係り受け関係の曖昧性を解消できる。

【図面の簡単な説明】

【図1】 本発明に係る第1の実施形態である連続音声認識装置のブロック図である。

【図2】 図1の連続音声認識装置81の音声認識動作を説明するために用いた第1と第2の2つの係り受け関係を有する一例文を示す図である。

【図3】 図1の連続音声認識装置81の音声認識動作の一例をスタック形式で示す図である。

【図4】 図1の連続音声認識装置81の音声認識動作の図3とは異なる例をスタック形式で示す図である。

【図5】 図2の例文の音声区間、ポーズ（無音区間等）及びポーズ単位（区切り単位）を示す図である。

【図6】 図1の連続音声認識装置81のLRパーザ5によって実行される音声認識処理を示すフローチャートである。

【図7】 本発明に係る第2の実施形態である連続音声認識装置82のブロック図である。

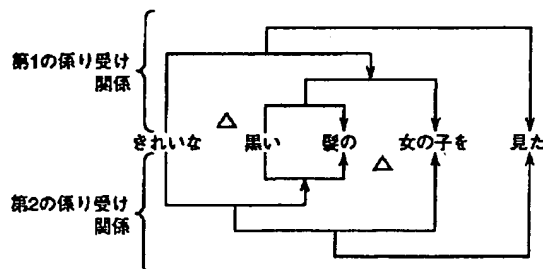
【図8】 図7の連続音声認識装置82のLRパーザ5によって実行される音声認識処理を示すフローチャートである。

【図9】 従来例の連続音声認識装置の音声認識動作をスタック形式で示す図である。

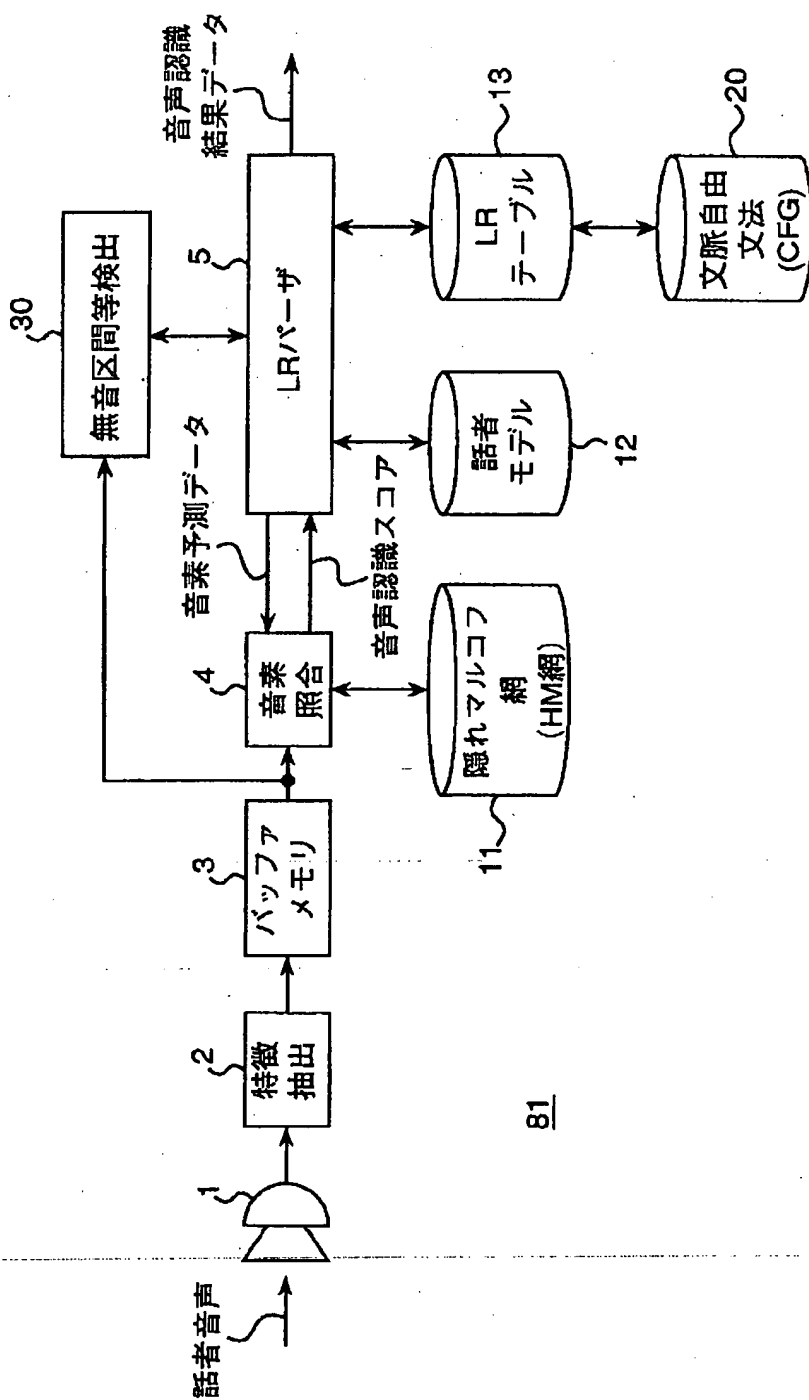
【符号の説明】

- 1…マイクロホン、
- 2…特徴抽出部、
- 3…バッファメモリ、
- 4…音素照合部、
- 5…LRパーザ、
- 11, 11a…隠れマルコフ網メモリ、
- 12…話者モデルメモリ、
- 13…LRテーブルメモリ、
- 20…文脈自由文法データベースメモリ、
- 30…無音区間等検出部。

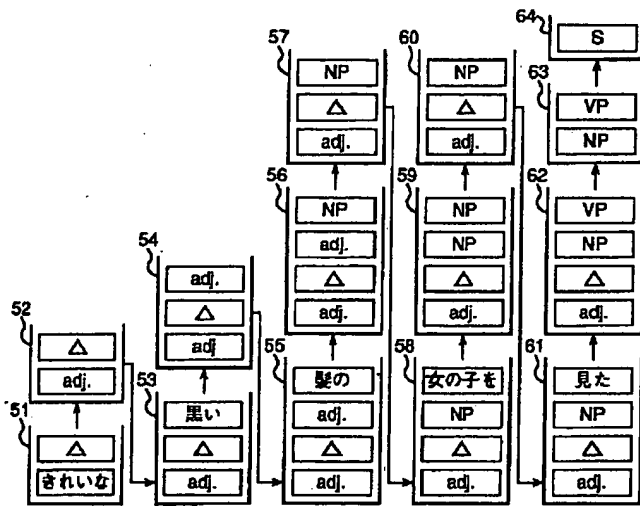
【図2】



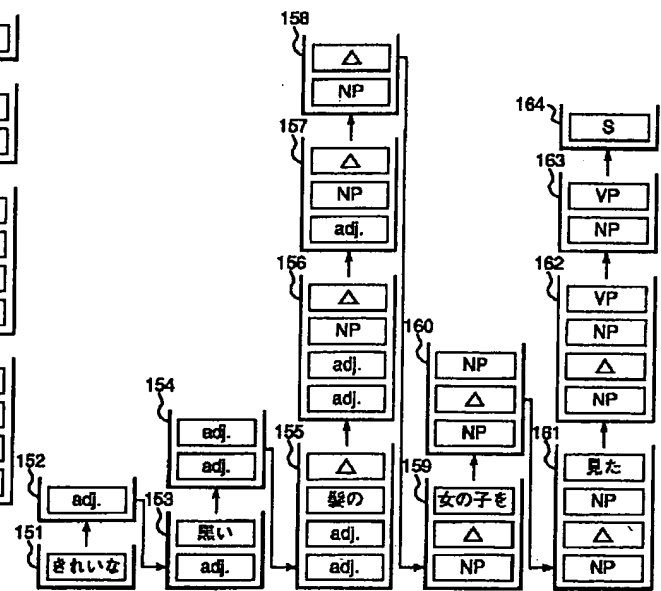
【図1】



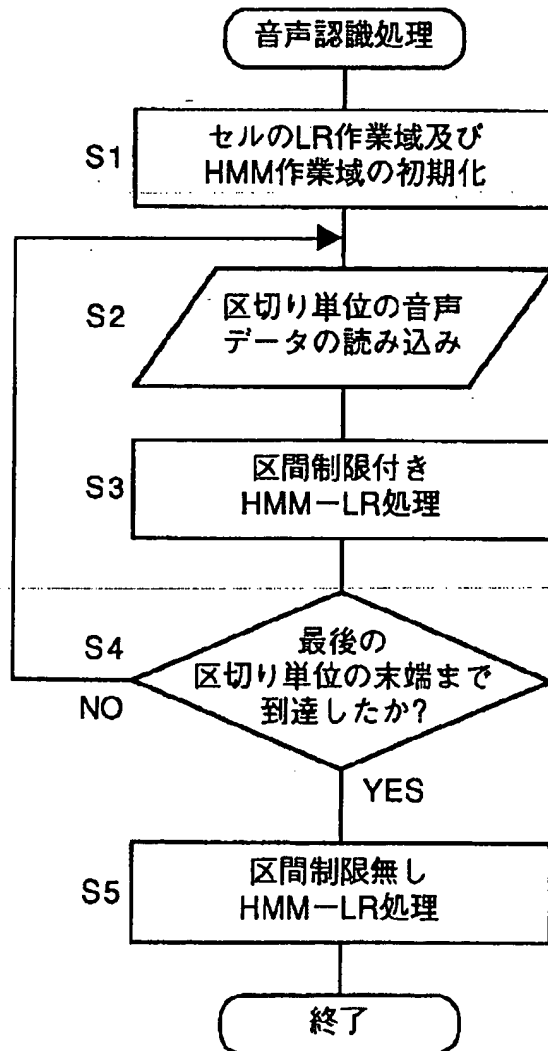
【図3】



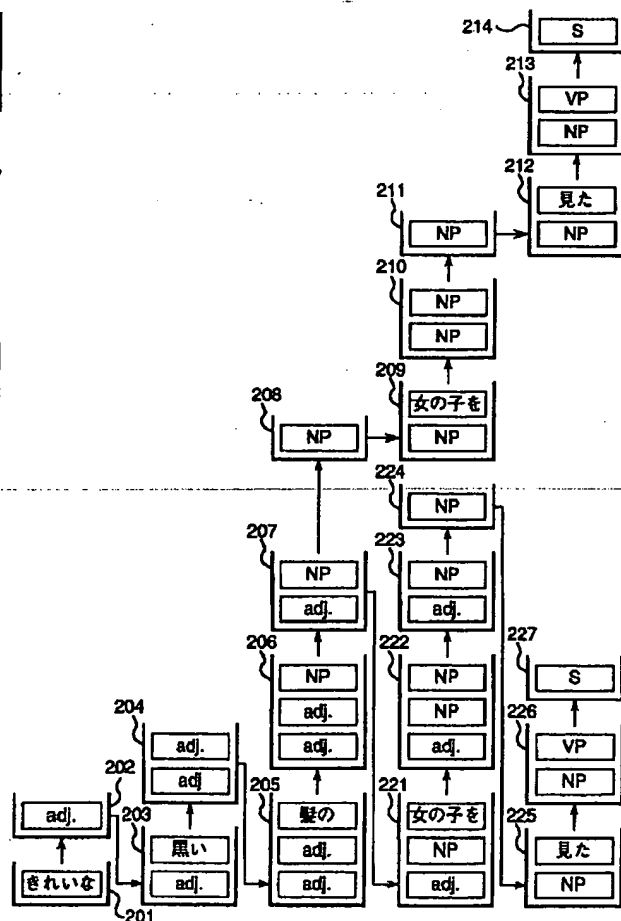
【図4】



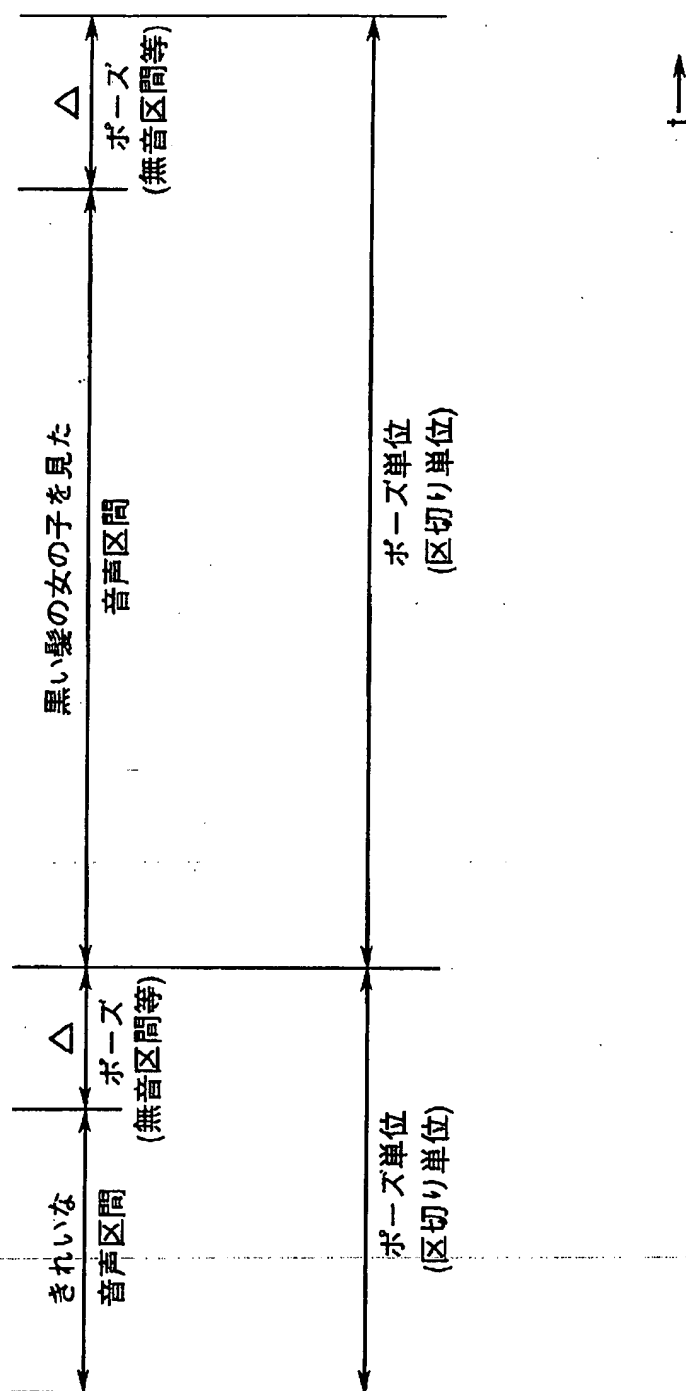
【図6】



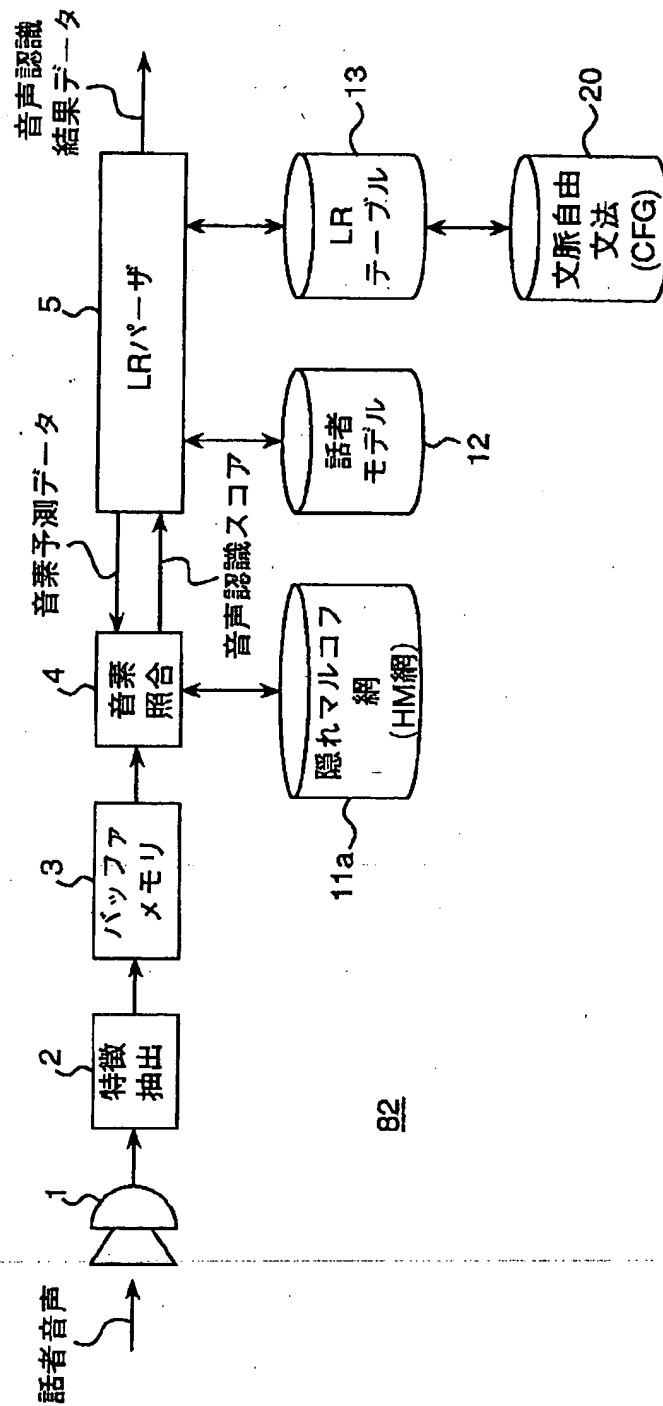
【図9】



【図5】



【図 7】



【図8】

